# Random Variables - Introduction

- Dealing with a summary variable is often easier than dealing with the original probability structure.
- Example: Opinion poll with "agree" and "disagree" responses.
- Define a variable $X$ as the number of people who agree.

## Definition of Random Variable

**Definition**
For a given random experiment with a sample space $S$, a function $X(\cdot)$ that assigns to each element $s$ in $S$ one and only one real number $X(s) = x$ is called a **random variable**.

**Examples of Random Variables**

| Experiment | Random Variable $X$ | Sample Space $\mathcal{X}$ |
|---|---|---|
| Toss two dice | Sum of the numbers | $\{2, 3, \ldots, 12\}$ |
| Toss a coin 25 times | Number of heads | $\{0, 1, \ldots, 25\}$ |
| Apply fertilizer to corn | Yield per acre | $[0, \infty)$ |

# Examples (cont'd)

**Example**
Tossing a fair coin once. Let $X$ denote the number of heads.

**Note:** A random variable will be denoted by an uppercase letter, e.g., $X$, and the realized values by corresponding lowercase letters, e.g., $x$.

**Example**
**(Distribution of a random variable)** Toss a fair coin 3 times. Let $X =$ number of heads in the three tosses. Find the distribution of $X$.

# Examples (cont'd)

**Solution:** Here $X$ takes the following values for each sample point in $S$.

| $s$ | {HHH} | {HHT}, {HTH}, {THH} | {TTH},{THT},{HTT} | {TTT} |
|---|---|---|---|---|
| $X$ | 3 | 2 | 1 | 0 |

So, the range for the random variable $X$ is $\mathcal{X} = \{0, 1, 2, 3\}$. Hence, the induced probability function on $\mathcal{X}$ is given by

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P_X(X = x)$ | $1/8$ | $3/8$ | $3/8$ | $1/8$ |

# Cumulative Distribution Function

**Definition**
The cumulative distribution function (cdf) of a random variable $X$, denoted by $F_X(x)$, is defined as $F_X(x) = P(X \leq x)$.

- $F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0$ and $F_X(\infty) = \lim_{x \to \infty} F_X(x) = 1$.
- $F_X(x)$ is nondecreasing.
- $F_X(x)$ is right-continuous.

# Ex: cdf

**Example**

Tossing three fair coins. Let $X$ :=number of heads. Then, the cdf of $X$ is

$$F_X(x) = \begin{cases} 0 & \text{if} \quad -\infty < x < 0 \\ 1/8 & \text{if} \quad 0 \leq x < 1 \\ 1/2 & \text{if} \quad 1 \leq x < 2 \\ 7/8 & \text{if} \quad 2 \leq x < 3 \\ 1 & \text{if} \quad 3 \leq x < \infty. \end{cases}$$

See Figure 3 below.

# Ex: discrete cdf

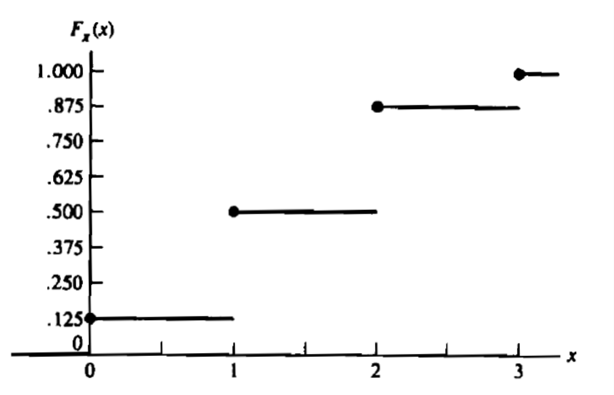


**Figure 1:** cdf plot for number of heads tossing three fair coins

# Ex: logistic cdf

**Example**
**(Continuous cdf)** The continuous function

$$F_X(x) = \frac{1}{1 + e^{-x}} \text{ for all } x$$

satisfies all the 3 properties of a cdf, and hence it is a cdf. It is actually known as the *logistic cdf*.

# Ex: logistic cdf

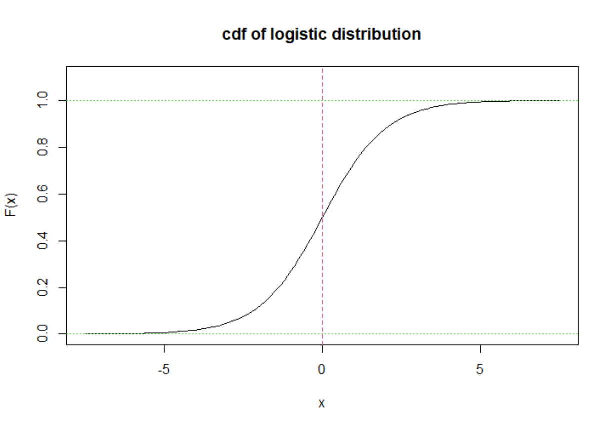


**Figure 2:** cdf curve for the logistic distribution

# Probability Density and Mass Functions

**Definition**
The probability mass function (pmf) of a discrete random variable $X$ is given by $f_X(x) = P(X = x)$ for all $x$.

**Definition**
The probability density function (pdf) of a continuous random variable $X$ is $f_X(x)$ which satisfies

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt.$$

**Equivalences and Theorems**

**Theorem**
*A function $f_X(x)$ is a pdf (pmf) of a random variable $X$ if and only if:*

1. $f_X(x) \geq 0$ *for all* $x$.
2. $\sum_x f_X(x) = 1$ *(discrete) or* $\int_{-\infty}^{\infty} f_X(x)dx = 1$ *(continuous).*

# How to get pmf or pdf from cdf

- If $X$ is discrete $f_X(a) = P(X = a)$. In particular, if $\mathcal{X}$ consists of $a_1 < a_2 < a_3 < \cdots$,

  $f_X(a_k) = P(X = a_k) = F_X(a_k) - F_X(a_{k-1})$ for all $a_k \in \mathcal{X}$ with $k > 1$

  and $f_X(a_1) = P(X = a_1) = F_X(a_1)$
- If $X$ is continuous, we get

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(t)dt,$$

then we have

$$\frac{d}{dx} F_X(x) = f_X(x),$$

which follows from the *Fundamental Theorem of Calculus*.

# Notation & Continuous Distribution Properties

**A note on notation:**

- $X$ has a distribution given by $F_X(x)$: $X \sim F_X$ or $X$ has pmf (or pdf) $f_X(x)$: $X \sim f_X$:
- $X \overset{d}{=} Y$ : $X$ and $Y$ have the same distribution ($X$ and $Y$ are identically distributed).

For a continuous random variable, $X$, with cdf $F_X(x)$, the following holds for all $a, b \in \mathbb{R}$ with $a \leq b$:

$$\begin{aligned} F_X(b) - F_X(a) &= P(a < X < b) = P(a \leq X < b) \\ &= P(a < X \leq b) = P(a \leq X \leq b) = \int_a^b f_X(x)dx, \end{aligned}$$

which also implies $P(X = c) = 0$ for any constant $c$.

# Continuous Distribution Properties

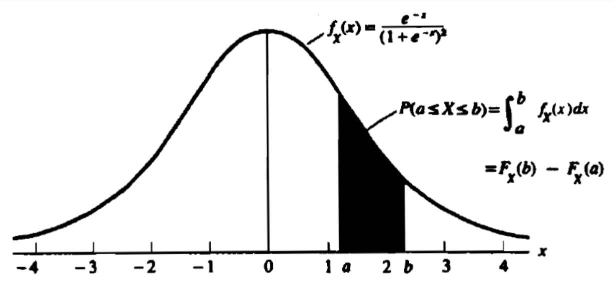


**Figure 3:** Illustration of probability as area under the pdf curve for logistic distribution

**Example**

Obtain the pdf of a logistic random variable with cdf

$$F_X(x) = \frac{1}{1 + e^{-x}}.$$

**Example**

Let $X$ denote the amount of space occupied by a (random) item placed in a 1-ft$^3$ packing container. Suppose the pdf of $X$ is given as:

$$f_X(x) = \begin{cases} kx^8(1 - x) & 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

(1) Find the value of $k$ that will make this a valid density function.

(2) Obtain the cdf of $X$.

(3) Find $P(0.2 \leq X \leq 0.8)$.

# Examples

**Solution:**

(1) $\int_0^1 kx^8(1-x)dx = 1$ implies that

$k\int_0^1 (x^8 - x^9)dx = k(x^9/9 - x^{10}/10)\Big|_{x=0}^{x=1} = 1$. Hence

$k(1/9 - 1/10) = 1 \quad \Rightarrow \quad k = 90.$

(2) Then $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$, so

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ 10x^9 - 9x^{10} & \text{if} \;\; 0 \leq x < 1 \\ 1 & \text{if} \;\; 1 \leq x \end{cases}$$

(3) (i) Using the pdf

$P(.2 \leq X \leq .8) = \int_{.2}^{.8} f_X(x)dx = \int_{.2}^{.8} 90x^8(1-x)dx = \ldots \approx .376.$

(ii) Using the cdf $P(.2 \leq X \leq .8) = F_X(.8) - F_X(.2) \approx .376.$

# Ex: Point Chosen in the Unit Circle

**Example**

Suppose we pick a point at random from the interior of a circle of radius 1 (and for the unit circle, origin = $(0, 0)$ is the center). Let $Z$ be the distance of the selected point from the origin. Then the sample space for the experiment is

$$S = \{z : 0 \leq z < 1\}.$$

(a) Find the cdf and the pdf of $Z$.

**Solution: Point Chosen in the Unit Circle**

- For $0 < z < 1$, event $\{Z \leq z\}$ is equivalent to the point lying in a circle of radius $z$.
- Let $C_z = \{(x, y) : x^2 + y^2 < z^2\}$.

# Conclusion

- Random variables help us summarize complex probability structures.
- They can be discrete or continuous (or mixed).
- The cumulative distribution function (cdf) and probability density/mass function (pdf/pmf) are essential concepts for understanding random variables.
- Equivalences like the one presented in the theorem help us identify valid pdfs/pmfs.

# Distributions of Functions of a Random Variable

- For a random variable $X$ with cumulative distribution function (cdf) $F_X(x)$, we're often interested in the behavior of functions of $X$, called transformations.
- In this section, we'll study the distribution of transformed random variables.
- Let's explore how different functions of $X$ impact the distribution of the resulting random variable $Y = g(X)$.

# Transformations of Random Variables

- If $X$ is a random variable with cdf $F_X(x)$, any function of $X$, denoted as $g(X)$, is also a random variable.
- Write $Y = g(X)$, and we can describe the probabilistic behavior of $Y$ in terms of $X$.
- The distribution of $Y$ depends on the functions $F_X$ and $g$.
- For any set $A$,

$$P(Y \in A) = P(g(X) \in A).$$

- The mapping $g$ can be one-to-one or onto, and its inverse $g^{-1}$ takes sets into sets.

# Discrete Case: Probability Mass Function (pmf)

- If $X$ is a discrete random variable with pmf $f_X(x)$, then $Y = g(X)$ is also discrete.
- The pmf of $Y$ is given by

$$f_Y(y) = P(Y = y) = \sum_{x:g(x)=y} f_X(x) \quad \text{for } y \in \mathcal{Y},$$

where $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$ and $\mathcal{X}$ is the support set of $X$.

# Example: Discrete Transformation

**Example**
Suppose $X$ takes values $i$ with probability $f_X(i) = 1/6$ for $i = 1, 2, 3, 4, 5, 6$. If $Y = (X - 3)^2$, find $\mathcal{Y}$ and $f_Y(y)$.

**Solution:** $\mathcal{Y} = \{0, 1, 4, 9\}$, and $f_Y(0) = f_X(3) = 1/6$, $f_Y(1) = f_X(2) + f_X(4) = 2/6$, $f_Y(4) = f_X(1) + f_X(5) = 2/6$, and $f_Y(9) = f_X(6) = 1/6$.

# Continuous Case: The cdf Technique

- If $X$ and $Y = g(X)$ are continuous random variables, we can find expressions for the cdf and pdf of $Y$ in terms of those of $X$.
- The cdf of $Y$ is given by

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y) = \int_{\{x \in \mathcal{X} : g(x) \leq y\}} f_X(x)dx.$$

- This technique is called *the cdf technique*.

# Ex: The cdf Technique

**Example**
**(Relation between Uniform and Exponential Distributions)**
Suppose $X$ has a uniform distribution on $(0, 1)$: $f_X(x) = 1$ if $0 < x < 1$ and 0 otherwise. Let $Y = -\log(X)$. Obtain the pdf and cdf of $Y$.

# Continuous Case: Probability Density Function (pdf)

- Recall that the pdf of $Y$ is obtained by differentiating the cdf with respect to $y$.

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

- From this, using the derivative rule for the inverse functions, one can obtain the pdf of $Y$ (provided that $g(x)$ is monotone).
- Thus, the pdf of $Y$ in terms of those of $X$ is

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| \quad \text{for } y \in \mathcal{Y},$$

where $\mathcal{Y} = \{y : y = g(x), x \in \mathcal{X}\}$.

# Ex: Continuous Transformation

**Example**

Suppose $X$ has a uniform distribution on $(0, 1)$: $f_X(x) = 1$ if $0 < x < 1$ and 0 otherwise. Obtain the pdf of $Y = -\log(X)$.

# Example: Continuous Transformation

**Example**

Let $X$ have pdf $f_X(x) = \begin{cases} (x+1)/2 & \text{for } -1 \le x \le 1, \\ 0 & \text{otherwise.} \end{cases}$ Find the pdf of $Y = X^2$.

**Solution:** $-1 \le X \le 1$ corresponds to $0 \le Y \le 1$. The cdf of $Y$ is

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx = \ldots = \sqrt{y} \quad \text{for } 0 < y < 1$$

So, $F_Y(y) = \begin{cases} 0 & \text{for } y \le 0, \\ \sqrt{y} & \text{for } 0 < y \le 1 \\ 1 & \text{for } y > 1 \end{cases}$

The pdf of $Y$ is $f_Y(y) = \frac{dF_Y(y)}{dy} = ... = \begin{cases} \frac{1}{2\sqrt{y}} & \text{for } 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases}$

# Conclusion

- Transformations of random variables are essential in probability and statistics to model real-world scenarios.
- The distribution of a transformed random variable depends on the cumulative distribution function of the original random variable and the transformation function.
- Discrete transformations have pmfs, while continuous transformations have pdfs derived using the cdf technique.
- Understanding these concepts is crucial for solving problems in various fields, including engineering, finance, and data science.

# Random Variables

**Definition**
A **random variable** is a numerical outcome of our experiment.

Random variables are fundamental in probability and statistics.

**Discrete Random Variables**

- In our dice example, the random variable $X$ could take on six values.
- Support of $X$ is a finite set.
- Discussed other examples of discrete random variables.

# Independent Random Variables

*Random variables X and Y are said to be* **independent** *if for any sets A and B, the events $\{X$ is in $A\}$ and $\{Y$ is in $B\}$ are independent, i.e. P(X is in A and Y is in B) = P(X is in A) P(Y is in B).*

Ex: Roll two dice, with X and Y denoting the number of dots on the blue and yellow dice. It is intuitively clear that the random variables X and Y not "affect" each other. If I know, say, that X = 6, that knowledge won't help me guess Y at all. For instance, the probability that Y = 2, knowing X=6, is still 1/6.

That is,

$$P(Y = 2|X = 6) = P(Y = 2)$$

which in turn implies

$$P(X = 6 \text{ and } Y = 2) = P(X = 6)P(Y = 2)$$

# Example: The Monty Hall Problem

**Introduction**
An illustration of how random variables can simplify the translation of a probability problem to mathematical terms.

**Background**

- Named after a TV game show host.
- Contestant chooses one of three doors.
- One door hides a car, the others hide goats.
- Contestant's goal: find the car.

# Example: The Monty Hall Problem

## The Twist

- Host knows where the car is.
- After the contestant chooses, the host opens a goat door.
- Should the contestant switch to the unopened door?

## Common Misconception

**Assumption**
Both unopened doors have a 1/2 chance of hiding the car.

**Reality**
Remaining door (unchosen and unopened) has a 2/3 chance!

# Example: The Monty Hall Problem

## Defining the Problem with Random Variables

- $C$: contestant's door choice (1, 2, or 3)
- $H$: host's door choice (1, 2, or 3) after contestant chooses
- $A$: door with the car

## Mathematical Formulation

Considering the case $C = 1,\ H = 2$:
Then the problem is to find the probability that the contestant should change her mind, i.e., the probability that the car is actually behind door 3:

$$P(A = 3 \mid C = 1,\ H = 2) = \frac{P(A = 3,\ C = 1,\ H = 2)}{P(C = 1,\ H = 2)} \tag{1}$$

# Example: The Monty Hall Problem

**Role of the Host**

The host's knowledge influences the problem's outcomes. The mathematical expression considering this role:

$$P(A = 3,\ C = 1)\ P(H = 2 \mid A = 3,\ C = 1) \tag{2}$$

**Paul Erdös's Mistake**

Even the famous mathematician Paul Erdős initially got it wrong. This showcases the importance of a structured mathematical approach.

# Expected Value

# Expected Value

**Overview**
Understanding the concept of expected value is central to probability and statistics.

**Generality of the Concept**

- The concepts and properties introduced apply to both discrete and continuous random variables.
- Properties extend to variance as well.

**"Expected Value":**

**A Misnomer, Not Always What We "Expect"**
The term "expected value" often does not align with intuition.

- Example: Expected heads in 1000 coin tosses is 500, but $P(H = 500)$ is around 0.025.
- Dice roll: Expected value is 3.5, but a die never shows a 3.5.

# Definition of Expected Value

**Expected Value of a Random Variable**
Expected value is the long-run average value of a random variable X. That is, the long-run average value as the experiment is repeated indefinitely.

$$\lim_{n \to \infty} \frac{X_1 + ... + X_n}{n} \tag{3}$$

Example: Rolling two dice and summing their faces. The expected sum is the long-run average of sum of the two dice.

# Existence of the Expected Value

- The definition assumes the limit exists.
- In practice, the limit exists if random variables have finite bounds, i.e., if values (of the rv) have finite upper and lower bounds.
- Real-world scenarios always adhere to this rule.
- We often refer to "the" expected value without the "if it exists" qualifier.

# Computation and Properties of Expected Value

# Computation and Properties of Expected Value

**Objective**
Understand the computational formula for the expected value of a discrete random variable.

**Coin Toss Experiment**

- Experiment: Toss 10 coins.
- Random variable $X$: Number of heads in 10 tosses.
- Observations:
  - $X_1 = 4$ (Four heads in first repetition)
  - $X_2 = 7$ (Seven heads in second repetition)
  - ... and so on.
- Intuitive long-run average of $X$: 5

Thus, $E(X) = 5$.

# Deriving the Formula

**Starting with Definition of Expected Value**

$$E(X) = \lim_{n\to\infty} \frac{X_1 + ... + X_n}{n}$$

**Rearranging with $K_{in}$ groups**

$$E(X) = \sum_{i=0}^{10} i \cdot \lim_{n\to\infty} \frac{K_{in}}{n}$$

- $K_{in}$: Number of times value $i$ occurs among $X_1, ..., X_n$.
- Group by number of heads. E.g.,
  $2 + 3 + 1 + 2 + 1 + 2 = 3 \times 2 + 2 \times 1 + 1 \times 3$.

# Expected Value Formula

- $\lim_{n\to\infty} \frac{K_{in}}{n}$ is the long-run fraction where $X = i$.
- This limit is essentially $P(X = i)$.

**Final Expected Value Formula**

$$E(X) = \sum_{i=0}^{10} i \cdot P(X = i)$$

In general,

**Property A: Expected Value of a Discrete Random Variable**

The expected value of a discrete random variable $X$ which takes values in the set $\mathcal{X}$ is:

$$E(X) = \sum_{i \in \mathcal{X}} i \cdot P(X = i) \tag{4}$$

**Insights on Expected Value**

- Expected value $E(X)$ is a weighted average.
- Weights $=$ Probabilities of the values.
- Some values of $X$ appear more frequently, influencing the average.
- Expected value $E(X)$ is constant.

For continuous random variables, the summation becomes an integral.

# Example: Coin Tosses

(a)

$$P(X = i) = \binom{10}{i} 0.5^i (1 - 0.5)^{10-i}$$

$$E(X) = \sum_{i=0}^{10} i \binom{10}{i} 0.5^i (1 - 0.5)^{10-i}$$

Result: $E(X) = 5$

(b) $N$: Number of coin tosses till we obtain a head

$$E(N) = \sum_{c=1}^{\infty} c \cdot \frac{1}{2^c} = 2$$

## Example: Dice

X: the number of dots we get in one roll of a die

S: sum of the dots on our yellow and blue die

$$E(X) = \sum_{c=1}^{6} c \cdot \frac{1}{6} = 3.5$$

$$E(S) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \cdots + 12 \cdot \frac{1}{36} = 7$$

Intuitive understanding: $E(X)$ represents the long-run average value.

# Dice Problem: Expanded Notebook

| notebook line | outcome | blue+yellow = 6? | S |
|---|---|---|---|
| 1 | blue 2, yellow 6 | No | 8 |
| 2 | blue 3, yellow 1 | No | 4 |
| 3 | blue 1, yellow 1 | No | 2 |
| 4 | blue 4, yellow 2 | Yes | 6 |
| 5 | blue 1, yellow 1 | No | 2 |
| 6 | blue 3, yellow 4 | No | 7 |
| 7 | blue 5, yellow 1 | Yes | 6 |
| 8 | blue 3, yellow 6 | No | 9 |
| 9 | blue 2, yellow 5 | No | 7 |

**Table 1:** Expanded Notebook for the Dice Problem

Note: $E(Y)$ for the yellow die will also be 3.5.

# Property B: Expected Value of a Sum & Properties C

**Property B:**

For random variables $U$ and $V$:

$$E(U + V) = E(U) + E(V) \tag{5}$$

Note: $U$ and $V$ do **not** need to be independent for this to hold.
This can be verified with the notebook analogy.

**Properties C**

- For any random variable $U$ and constant $a$: $E(aU) = aEU$
- For any constant $b$: $E(b) = b$

**Intuition:** If $X$ is constant with value 3, then $EX = 3$.

# Special Cases

- For $X$ and $Y$: $E(aX + bY) = aEX + bEY$
- Combining properties, we get:

$$E(aX + b) = aEX + b$$

**Example: Temperature Conversion**

Say $U$ is temperature in Celsius. Then the temperature in Fahrenheit is $W = \frac{9}{5}U + 32$. Using our property, we can find its expected value of the temperature in Fahrenheit, if the expected value in Celsius is available.

# Property D

If $E[g(X)]$ exists, then

$$E[g(X)] = \sum_{c \in \mathcal{X}} g(c) \cdot P(X = c)$$

where the sum is over all values $c$ that $X$ can take.

**Example: (Die Roll)** Let's find $E(\sqrt{X})$ where $X$ is the number of dots on a die roll.

$$E(\sqrt{X}) = \sum_{i=1}^{6} \sqrt{i} \cdot \frac{1}{6} \approx 1.805$$

# Property E

If $U$ and $V$ are independent:

$$E(UV) = EU \cdot EV$$

Equation $E(UV) = EU \cdot EV$ lacks an easy notebook illustration. Refer to Appendix Section for detailed proof.

Example: Dice with blue and yellow dots, $D = XY$.

$$E(D) = E(XY) = EX \cdot EY = 3.5^2 = 12.25.$$

# Importance of $E[g(X)]$

$E[g(X)]$ is crucial and will be frequently used.

**Mailing Tubes Concept**

- Properties of expected value are central to understanding.
- Equations like $E[g(X)]$ are "mailing tubes".
- Recognize scenarios to apply these properties.

Remember to always utilize the "mailing tubes" throughout your studies and work.

# Bus Ridership: Expected Values

- Find the expected value of $N_1$, the number of passengers on the bus as it leaves the first stop.
- Extend the concept to find $E[N_2]$.

**Given Data**

- Bus arrives empty at the first stop.
- $N_1 = B_1$, where $B_1$ is the number who board at the first stop.
- Support of $N_1$ (equivalently $B_1$): 0, 1, and 2 (i.e., $\mathcal{N}_1 = \{0, 1, 2\}$).
- Probabilities: $P(B_1 = 0) = 0.5$, $P(B_1 = 1) = 0.4$, and $P(B_1 = 2) = 0.1$.

# Calculation of $E[N_1]$

Using the formula:

$$E[N_1] = \sum_{i \in \mathcal{N}_1} i \times P(N_1 = i)$$

$$E[N_1] = 0(0.5) + 1(0.4) + 2(0.1) = 0.6$$

**Interpretation of $E[N_1]$:** On observing the bus over many days, on average, it will leave the first stop with 0.6 passengers.

**Extension to $E[N_2]$**

- Support of $N_2$: $\{0, 1, 2, 3, 4\}$.
- Need to find $P(N_2 = i)$ for $i = 0, 1, 2, 3, 4$.
- Known: $P(N_2 = 0) = 0.292$.
- Other probabilities can be determined similarly as above.

# Expected Values via Simulation

When the expected values $EX$ are too intricate to determine analytically, one can use simulation as an alternative. By understanding the expected value as a long-run average, the approach is straightforward:

- Execute `nreps` replications of the experiment.
- For each run, record the value of $X$.
- Calculate the average over `nreps` values.

To illustrate this, a modified version of the code from Section 2.4 is provided, aiming to estimate the expected number of passengers on a bus as it departs from the tenth stop.

# Expected Values via Simulation

```
nreps <- 10000
nstops <- 10
total <- 0
for (i in 1: nreps ) {
passengers <- 0
for (j in 1: nstops ) {
if ( passengers > 0)
for (k in 1: passengers )
if ( runif (1) < 0.2)
passengers <- passengers - 1
newpass <- sample (0:2 ,1 , prob =c (0.5 ,0.4 ,0.1))
passengers <- passengers + newpass
}
total <- total + passengers
}
print ( total / nreps )
```

# Expected Value for Continuous RVs

The *expected value* or *mean* of a random variable $g(X)$, denoted by $E[g(X)]$, is

$$E[g(X)] = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x) dx & \text{if } X \text{ is continuous,} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) P(X = x) & \text{if } X \text{ is discrete,} \end{cases}$$

provided that the integral or sum exists.

## Mean of the Exponential Distribution

Suppose $X$ has an Exponential($\lambda$) distribution, that is, it has pdf given by

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

for $\lambda > 0$. Find the mean of $X$.

**Solution**

$$E[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = -xe^{-\lambda x}\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx =$$

$$-\frac{1}{\lambda}e^{-\lambda x}\Big|_0^\infty = \frac{1}{\lambda} \text{ (integration by parts)}$$

# Example: Relation between Uniform and Exponential Distributions - II

Let $X$ have a Uniform$(0, 1)$ distribution, i.e., its pdf is

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Define $Y = g(X) = -\log X$ and find $E[Y]$.

# Solution

**Solution 1:** Using the density of

$$E[Y] = E[g(X)] = E[-\log X] = \int_0^1 -\log x dx =$$

$$-x\log x\Big|_0^1 + \int_0^1 -x\frac{d(\log x)}{dx}dx = 1.$$

**Solution 2:**

Using the density of $Y$. Recall

$$f_Y(y) = \frac{d}{dy}\left(1 - e^{-y}\right) = e^{-y}$$

for $0 < y < \infty$.

This is the special case of exponential distribution with $\lambda = 1$.
Thus, $EY = 1/\lambda = 1$.

## Central Tendency: Mean vs. Median

- Expected value is a measure of central tendency.
- Other measures: Median (halfway point of distribution).
- Mean has historical significance in probability and statistics.

# Misconceptions about Expected Value

- Not always the value we truly “expect”.
- Example: Expected number of dots on a die is 3.5, but this is impossible in reality.

**Limitations of the Mean**

- Can be skewed by outliers, e.g., if Bill Gates moved to a town.
- Mean might not capture the true central tendency.
- The significance of mean in real-life scenarios might be ambiguous.

# When Mean Makes Sense

- Useful when interested in totals (e.g., total defects in a batch).
- In many scenarios (e.g., describing wealth or exam scores), the total is not directly relevant.
- The median might be more representative in such cases.

**Mean in Business: Casinos and Insurance Companies**

- Mean has tangible significance for casinos and insurance companies.
- Useful for predicting totals over long runs (e.g., total payouts).
- Helps businesses plan and set prices or premiums.

# Introduction

# Variance

As in Section on Expectation (Ch 3 in Matloff), the concepts and properties introduced in this section form the very core of probability and statistics.

**Note:** Except for some specific calculations, these apply to both discrete and continuous random variables.

# Definition

While the expected value tells us the long-run average of a random variable, we also need a measure of its variability.

In other words, we want a measure of **dispersion** or **spread**. The classical measure of spread is **variance**.

**Variance Definition:** For a random variable $U$ for which the expected values exist, the **variance** of $U$ is defined to be:

$$Var(U) = E[(U - EU)^2] \quad (1)$$

## Die Example

For $X$ in the die example:

$$Var(X) = E[(X - 3.5)^2] \tag{2}$$

Here, $W = (X - 3.5)^2$ is a function of $X$. We find the expected value of this new random variable $W$.

**Notebook View**

| line | X | W |
|---|---|---|
| 1 | 2 | 2.25 |
| 2 | 5 | 2.25 |
| 3 | 6 | 6.25 |
| 4 | 3 | 0.25 |
| 5 | 5 | 2.25 |
| 6 | 1 | 6.25 |

# Calculating Variance

To evaluate, apply:

$$Var(X) = \sum_{c=1}^{6} (c - 3.5)^2 \cdot \frac{1}{6} = 2.92$$

Variance gives us a measure of dispersion.

**Intuition**

In the expression $Var(U) = E[(U - EU)^2]$:

- If values of $U$ are clustered near its mean, then $(U - EU)^2$ will usually be small.
- Then variance of $U$ will be small.
- If $U$ varies widely, then its variance will be large.

# Properties of Variance

# Property F

$$Var(U) = E(U^2) - (EU)^2 \quad (3)$$

The term $E(U^2)$ is evaluated using Property D of expectation.

**Derivation of Property F**

Using properties of expected values:

$$\begin{aligned} Var(U) &= E[(U - EU)^2] \\ &= E[U^2 - 2EU \cdot U + (EU)^2] \text{ (algebra)} \\ &= E(U^2) + E(-2EU \cdot U) + E[(EU)^2] \text{ (property B)} \\ &= E(U^2) - 2EU \cdot EU + (EU)^2 \text{ (properties C)} \\ &= E(U^2) - (EU)^2 \end{aligned}$$

**Note on Property F:** Remember, (3) is a shortcut formula for finding the variance, not the *definition* of variance.

## Example using Property F

If $X$ is the number of dots which come up when we roll a die:

$$Var(X) = E(X^2) - (EX)^2 \tag{4}$$

From Property D of expectation

$$E(X^2) = \sum_{i=1}^{6} i^2 \cdot \frac{1}{6} = \frac{91}{6} \tag{5}$$

Thus,

$$Var(X) = E(X^2) - (EX)^2 = \frac{91}{6} - 3.5^2 \approx 2.92$$

# Property G

An important behavior of variance is:

$$Var(cU) = c^2 Var(U) \tag{6}$$

This means: If we multiply a random variable by $c$, its variance multiplies by $c^2$.

**Proof of Property G:** Defining $V = cU$:

$$\begin{aligned} Var(V) &= E[(V - EV)^2] \text{ (def.)} \\ &= E\{[cU - E(cU)]^2\} \text{ (subst.)} \\ &= E\{[cU - cEU]^2\} \text{ (property C)} \\ &= E\{c^2[U - EU]^2\} \text{ (algebra)} \\ &= c^2 E\{[U - EU]^2\} \text{ (property C)} \\ &= c^2 Var(U) \text{ (def.)} \end{aligned}$$

# Property H

Shifting data over by a constant does not change the amount of variation in them:

$$Var(U + d) = Var(U) \tag{7}$$

for any constant $d$.

**Variance of a Constant:** Intuitively, the variance of a constant is 0 — after all, it never varies!

Formally:

$$Var(c) = E(c^2) - [E(c)]^2 = c^2 - c^2 = 0 \tag{8}$$

**Standard Deviation:** The square root of the variance is called the **standard deviation**: $SD(X) = \sqrt{Var(X)}$.

# An Alternative Measure of Dispersion

Variance is used historically and mathematically, not because it's the most meaningful measure (of spread).

- Squaring exaggerates larger differences.
- A more natural measure: **mean absolute deviation** (MAD), $E(|U - EU|)$.
- MAD is less mathematically tractable.

The choice of variance allows for powerful mathematical derivations (e.g., Pythagorean Theorem in abstract vector spaces).

# Variance of Sum of Independent RVs

If $U$ and $V$ are independent,

$$Var(U+V) = Var(U) + Var(V) \tag{9}$$

Generalizing (9), for constants $a_1, ..., a_k$ and independent random variables $X_1, ..., X_k$, form the new random variable $a_1X_1 + ... + a_kX_k$. Then

$$Var(a_1X_1 + ... + a_kX_k) = \sum_{i=1}^{k} a_i^2 Var(X_i) \tag{10}$$

# Importance of Variance

## Importance of Variance Properties

- The properties of variance are crucial for understanding the rest of the content.
- Recognize settings where they are applicable. Think of a property like (9) and check for independence.

## Central Importance of the Concept of Variance

- The mean is a fundamental descriptor of a random variable.
- Variance is of central importance.
- Used constantly throughout subsequent discussions.
- Next: A quantitative look at variance as a measure of dispersion.

## Intuition Regarding the Size of Var(X)

The variance of a random variable $X$ is a measure of dispersion. But, how do we quantify its magnitude?

**Chebychev's Inequality:** This inequality provides concrete meaning to the concept of variance/standard deviation:

$$P(|X-\mu| \geq c\sigma) \leq \frac{1}{c^2} \tag{11}$$

- For instance, $X$ strays more than 3 standard deviations from its mean at most only $1/9$ of the time.
- Used in grading schemes:
  "An A grade is 1.5 standard deviations above the mean" and
  "A C grade is 1.5 standard deviations below the mean"
  (here $c = 1.5$).
- Proof of the inequality provided later.

# The Coefficient of Variation

- Reflect on the magnitude of variance.
- E.g., if the price of a widget hovers around a $1 million, but the variation around that figure is only about a dollar, there is essentially no variation. But a variation of about a dollar in the price of a hamburger would be a lot.
- Relate size of $SD(X)$ to $E(X)$ for context.
- Define the coefficient of variation:

$$\text{coef. of var.} = \frac{SD(X)}{EX} = \frac{\sqrt{Var(X)}}{EX}$$

- A scale-free measure to judge the size of variance.

# A Useful Fact

For a random variable $X$:

$$g(c) = E[(X - c)^2] \tag{12}$$

The function $g(c)$ maps a real number $c$ to a real output. What value of $c$ minimizes $g(c)$?

Using the properties of expected value:

$$g(c) = E(X^2) - 2cEX + c^2 \tag{13}$$

Differentiate with respect to $c$ and set to 0 to find:

$$c = EX$$

# Optimal Guessing

- Consider guessing people's weights without any prior information.
- Initial inclination: use the mean weight as your guess.
- When you measure the error in your guess using "mean squared error":

  $$E[(X-c)^2]$$

  $c = EX$ minimizes the error.
- This confirms that the optimal guess is the mean weight.

# Conclusion and Alternate Consideration

- Plugging $c = EX$ into $g(c)$ shows the minimum value is $E(X - EX)^2$, i.e., $Var(X)$.
- Think of this in terms of long-run average squared error.
- Alternative: Minimize average absolute error:

$$E(|X - c|)$$

- The optimal $c$ for this is the **median** weight.

# Indicator Random Variables, and Their Means and Variances

# Definition: Indicator Random Variable

**Definition**
A random variable that has the value 1 or 0, depending on whether a specified event occurs or not, is called an **indicator random variable** for that event.

**Handy Facts**

- If $X$ is an indicator random variable for event $A$ and $p = P(A)$, then:

$$E(X) = p \tag{14}$$

$$Var(X) = p(1 - p) \tag{15}$$

- For example: $EX = P(X = 1) = P(A) = p$.

## Application Example

Consider three coins:

- Coin A has $P(heads) = 0.6$
- Coin B (fair) has $P(heads) = 0.5$
- Coin C has $P(heads) = 0.2$

Toss each once, recording heads as $X$, $Y$, and $Z$ respectively.
$W = X + Y + Z$ is the total number of heads.

**Find $P(W = 1)$ and $Var(W)$:** To find $P(W = 1)$:

$$\begin{aligned} P(W = 1) &= P(X = 1, Y = 0, Z = 0) + \dots \\ &= 0.6 \times 0.5 \times 0.8 + 0.4 \times 0.5 \times 0.8 + 0.4 \times 0.5 \times 0.2 \\ &= 0.44 \end{aligned}$$

To find $Var(W)$ using indicator random variables:

$$Var(W) = 0.6 \times 0.4 + 0.5 \times 0.5 + 0.2 \times 0.8 = 0.65$$

# Example: Return Time for Library Books, Version I

Suppose at a public library:

- Patrons return books exactly 7 days after borrowing.
- Returning to a different branch adds 2 days delay.
- 50% return their books to a "foreign" library.

Find $Var(T)$, where $T$ is the time (either 7 or 9 days) for a book to come back.

**Solution:**

$$T = 7 + 2I$$

where $I$ indicates if the book is returned to a "foreign" branch.

$$Var(T) = Var(7 + 2I) = 4Var(I) = 4 \times 0.5 \times (1 - 0.5) = 1.0$$

# Example: Return Time for Library Books, Version II

- Borrowers return books after 4, 5, 6 or 7 days with probabilities 0.1, 0.2, 0.3, 0.4.
- 50% return their books to a "foreign" branch (causing 2-day delay).
- Library is open 7 days a week.
- Suppose you wish to borrow a certain book, and inquire at the library near the close of business on Monday.
- You are told that it had been checked out the previous Thursday.
- Assume that no one else is waiting for the book, you check every evening, and a borrower returning to a foreign branch is independent of his/her return day.
- Find the probability of waiting until Wednesday evening.

# Example: Return Time for Library Books, Version II

**Solution** Let $B$: the time (# of days) needed for the book to arrive back at its home branch,
$R$: the amount of time it takes for borrowers to return books, and define $I$ as before. Note that $B = R + 2\,I$. Then

$$\begin{aligned} P(B = 6 \mid B > 4) &= \frac{P(B = 6 \text{ and } B > 4)}{P(B > 4)} \\ &= \frac{P(B = 6)}{1 - P(B \leq 4)} = \frac{P(B = 6)}{1 - P(B = 4)} \end{aligned}$$

So, $B = 6$ occurs when "R=6 and I=0" or "R=4 and I=1". Thus,

$$\begin{aligned} P(B = 6 \mid B > 4) &= \frac{P(\text{``}R = 6 \text{ and } I = 0\text{'' or ``}R = 4 \text{ and } I = 1\text{''})}{1 - P(R = 4 \text{ and } I = 0)} \\ &= \frac{0.3 \cdot 0.5 + 0.1 \cdot 0.5}{1 - 0.1 \cdot 0.5} \\ &= \frac{4}{19} \approx 0.21 \end{aligned}$$

# Simulation Check

```
libsim <- function(nreps) {
   prt <- sample(c(4,5,6,7),nreps,replace=T,
                 prob=c(0.1,0.2,0.3,0.4))
   i <- sample(c(0,1),nreps,replace=T)
   b <- prt + 2*i
   x <- cbind(prt,i,b)
   bgt4 <- x[b > 4,]
   mean(bgt4[,3] == 6)
}
```

Use *R*'s vector operations for easier coding and faster running.